

PENYUSUNAN INSTRUMEN TES TENGAH SEMESTER GENAP FISIKA X SMA UNTUK KELAS X SMA

¹⁾Dian Wahyu Nur Ivanty, ²⁾Nonoh Siti Aminah, ³⁾Elvin YuslianaEkawati

Fakultas Keguruan dan Ilmu Pendidikan
Universitas Sebelas Maret

Jl. Ir. Sutami 36 A, Surakarta, email : dianwahyu31@yahoo.com

ABSTRACT

The aim of this research was to apprehend the characteristics of physics mid semester test instrument which has been arranged. Research method used by descriptive method. Procedures used are adopt the procedure of arranging the test by Djemari Mardapi. Its use qualitative analysis technique which consists of aspect materials, construction, and language, and quantitative analysis using the program MicroCat Iteman version 3.00 to determine difficulty level, determinant indeks and effectiveness of detractors. The results of the test arrangement was an objective test in multiple choices form with five answer choices of 40 items. Time allocation of test doing was 90 minutes. Characteristics of final arranged test are: 1) difficulty level: 42,5% medium category and 57,5% difficult category; 2) determinant indeks: 5% excellent category, 35% good category, 25% satisfactory category, 27,5% poor category dan 7,5% very poor; 3)effectiveness of detractors: 95% four distractors functioned, dan 5% tri distractors functioned. The final result is about 60% received and 40% revised.

Key word: Test arrangement, difficulty level, determinant indeks and effectiveness of detractors

PENDAHULUAN

Menurut Nomor 14 tahun 2005 tentang Guru dan Dosen dan lebih lanjut dijelaskan dalam Peraturan Pemerintah Nomor 74 tahun 2008 tentang Guru salah satu tugas guru adalah melakukan evaluasi dan penilaian. Salah satu tahap penting dalam proses penilaian adalah pengumpulan informasi. Dalam penilaian pendidikan, informasi yang dikumpulkan umumnya berupa hasil belajar siswa, baik yang sifatnya pengetahuan ketrampilan maupun sikap. Lebih khusus lagi untuk keperluan pengambilan keputusan dalam hubungannya dengan proses belajar mengajar di kelas, biasanya digunakan informasi hasil belajar yang berupa pengetahuan. Alat ukur hasil belajar siswa yang sifatnya pengetahuan umumnya adalah tes (uji) (Cartono dan Utari, 2006:23). Sehingga keberhasilan kegiatan evaluasi hasil belajar di sekolah sangat tergantung pada kualitas instrumen tes yang digunakan. Dengan demikian, kemampuan guru dalam menyusun instrumen tes yang baik merupakan salah satu faktor yang berpengaruh terhadap keberhasilan kegiatan evaluasi di sekolah.

Pada kenyataan di lapangan masih terdapat guru yang mengalami kesulitan dalam menyusun soal, menganalisis serta menginterpretasikannya. Meskipun, jika dilihat secara umum guru dengan pengalamannya mengajar serta pengetahuannya terhadap karakteristik peserta didiknya memiliki kemampuan untuk menyusun soal. Namun, selama ini untuk pelajaran fisika guru relatif lebih sering melakukan penyusunan tes essay sehingga ketika melakukan penyusunan soal pilihan ganda mengalami kesulitan.

Kendala lain yang dimiliki guru adalah dalam hal analisis butir soal. Selama ini karena keterbatasan waktu dan kurangnya sosialisasi, banyak guru yang tidak melakukan

prosedur analisis butir soal. Anastari dan Untari (1997) menjelaskan bahwa tujuan utama analisis butir soal adalah untuk mengidentifikasi kekurangan-kekurangan dalam tes atau dalam pembelajaran (Depdiknas, 2008:1). Dengan alasan tersebut maka kegiatan analisis butir soal sangat diperlukan untuk meningkatkan kualitas instrumen tes yang dibuat guru.

Uraian diatas memberikan gambaran betapa pentingnya melakukan penyusunan tes fisika untuk evaluasi hasil belajar siswa. Berdasarkan dari alasan tersebut, maka penelitian ini berjudul **Penyusunan Instrumen Tes Tengah Semester Genap Fisika untuk Kelas X SMA.**

Rumusan masalah dari penelitian ini adalah Bagaimanakah karakteristik instrumen tes tengah semester genap fisika untuk kelas X SMA yang disusun?, berdasarkan rumusan masalah tersebut maka tujuan penelitian ini adalah untuk mengetahui karakteristik instrumen tes tengah semester genap fisika untuk kelas X SMA yang disusun.

KAJIAN TEORI

Menurut Griffin dan Nix yang dikutip oleh Djemari Mardapi (2008:1) “pengukuran, asesmen dan evaluasi adalah hirarki”. Evaluasi didahului dengan penilaian (*assessment*) sedangkan penilaian didahului dengan pengukuran. Pengukuran membandingkan hasil pengamatan dengan kriteria, asesmen menjelaskan dan menafsirkan hasil pengukuran, sedangkan evaluasi adalah penetapan nilai atau implikasi suatu perilaku.

Rammers, Gage dan Rummel (1960:7) berpendapat bahwa “*Measurement refers to observations that can be expressed quantitatively and answers the question “how much”*”. Pernyataan tersebut berarti pengukuran berkenaan dengan pengamatan yang dapat dinyatakan secara kualitatif dan dapat menjawab pertanyaan seberapa banyak. Penilaian (*Assesment*) menurut Gronlund & Linn (1995:5), mencakup berbagai prosedur yang digunakan untuk mendapatkan informasi tentang belajar siswa (melalui observasi, skala sikap atau proyek, tes tertulis) dan sebagai bahan pertimbangan tentang kemajuan belajar siswa. Lebih lanjut Djemari Mardapi mendefinisikan evaluasi sebagai proses mengumpulkan informasi untuk mengetahui pencapaian belajar kelas atau kelompok yang dapat dimanfaatkan untuk meningkatkan kualitas proses belajar mengajar (2008:9).

Secara umum teknik penilaian dapat digolongkan menjadi dua yaitu teknik tes dan nontes. Teknik tes umumnya digunakan untuk menilai kemampuan siswa yang mencakup pengetahuan dan ketrampilan sebagai hasil belajar, bakat khusus dan bakat umum. Sementara untuk teknik nontes umumnya digunakan untuk menilai karakteristik-karakteristik lain dari siswa misalnya minat, sikap dan kepribadian (Cartono dan Utari, 2006: 45).

Nana Sudjana (2005:35) berpendapat bahwa, “Tes sebagai alat penilai adalah pertanyaan-pertanyaan yang diberikan kepada siswa untuk mendapat jawaban dari siswa dalam bentuk lisan (tes lisan), dalam bentuk tulisan (tes tulisan) atau dalam bentuk perbuatan (tes tindakan).”

Dari segi bentuk tes dibagi menjadi tes lisan dimana tester di dalam mengajukan pertanyaan-pertanyaan atau soalnya dilakukan secara lisan, dan *testee* memberikan jawaban secara lisan pula dan tes tertulis dimana dalam mengajukan butir-butir pertanyaan atau soalnya dilakukan secara tertulis dan *testee* memberikan jawabannya juga secara tertulis (Anas Sudijono, 2005:75). Salah satu jenis tes tertulis adalah tes objektif yang terdiri dari tes benar salah, pilihan ganda, menjodohkan dan jawaban singkat (Suharsimi Arikunto, 2011:164).

Menurut Noer Munajir (1981) dalam Chalib Thoha (1994:71) mendefinisikan tes pilihan ganda sebagai tes objektif dimana masing-masing item disediakan lebih dari dua kemungkinan jawaban, dan hanya satu dari pilihan tersebut yang benar atau paling benar. Tes pilihan ganda terdiri atas bagian keterangan (stem) dan bagian kemungkinan jawaban alternatif (options). Kemungkinan jawaban (options) terdiri atas satu jawaban yang benar dan beberapa pengecoh (distractor)” (Anas Sudijono, 2005:118).

Tes objektif pilihan ganda memiliki beberapa kebaikan yang tidak dimiliki tes jenis lainnya, yaitu: 1) Tes objektif pilihan ganda bentuknya lebih representatif dalam hal mencakup dan mewakili materi pembelajaran. 2) Tes objektif pilihan ganda lebih memungkinkan bagi *tester* untuk bertindak lebih objektif, baik dalam, mengoreksi lembar jawaban soal, menentukan bobot skor, maupun dalam menentukan nilai akhirnya. 3) Butir-butir soal jauh lebih mudah dianalisis, baik analisis dari segi tingkat kesukaran, daya pembeda, maupun reliabilitasnya. 4) Dari segi efisiensi waktu, mengoreksi tes objektif pilihan ganda jauh lebih cepat.

Ada sembilan langkah yang perlu ditempuh dalam mengembangkan tes hasil atau prestasi belajar (Djemari Mardapi, 2008), yaitu: a) menyusun spesifikasi tes, b) menulis soal tes, c) menelaah soal tes atau analisis secara kualitatif, d) melakukan ujicoba tes, e) menganalisis butir soal atau analisis secara kuantitatif, f) memperbaiki tes, g) merakit tes, h) melaksanakan tes, i) menafsirkan hasil tes .

Analisis tes dilakukan dari segi kualitatif dan kuantitatif. Analisis secara kualitatif dilakukan oleh ahli, meliputi telaah terhadap aspek konstruksi, aspek materi dan bahasa. Analisis kuantitatif dilakukan dengan pendekatan teori tes klasik. Menurut Badrun Kertowirangan analisis kuantitatif menurut pendekatan teori tes klasik menghasilkan karakteristik yang meliputi: kesukaran (p), daya pembeda (d), dan keefektifan distraktor. Selain itu dapat pula diketahui reliabilitas soal tes dan kesalahan baku pengukuran. Analisis taraf kesukaran, daya beda dan keefektifan distraktor dilakukan melalui analisis tiap butir soal, sedangkan reliabilitas dan kesalahan pengukuran baku dilihat dengan cara analisis tes secara keseluruhan (2009:13).

Teknik analisis reliabilitas menggunakan teknik formula *Kuder-Richardson-20* atau KR-20, karena penskoran tes fisika yang dikembangkan berbentuk dikotomi. Untuk mengetahui reliabilitas butir soal menggunakan rumus alpha KR-20 adalah sebagai berikut:

$$KR - 20 = \left(\frac{k}{k-1} \right) \left(1 - \frac{\sum p(1-p)}{SD^2} \right) \quad (\text{Depdiknas, 2008:18})$$

Di mana p (proporsi subjek yang menjawab item benar), k (banyak item) dan SD (standar deviasi dari tes). Kriteria reliabilitas:

$0,00 \leq r \leq 0,19$: reliabilitas sangat rendah

$0,20 \leq r \leq 0,39$: reliabilitas rendah

$0,40 \leq r \leq 0,59$: reliabilitas cukup

$0,60 \leq r \leq 0,79$: reliabilitas tinggi

$0,80 \leq r \leq 1,00$: reliabilitas sangat tinggi (Suharsimi Arikunto, 2005 : 100)

Analisis kuantitatif yang lain adalah analisis mengenai taraf kesukaran. Taraf kesukaran tes adalah kemampuan tes tersebut dalam menjaring banyaknya subyek peserta tes yang dapat mengerjakan dengan betul. Rumus yang digunakan untuk menghitung indeks kesukaran adalah :

$$\text{Tingkat Kesukaran}(P) = \frac{\text{jumlah siswa yang menjawab benar butir soal}}{\text{jumlah siswa yang mengikuti tes}} \quad (\text{Depdiknas, 2008:12})$$

Klasifikasi soal :

- $1,00 \leq p \leq 0,29$: soal kategori sukar
- $0,30 \leq p \leq 0,69$: soal kategori sedang
- $0,70 \leq p \leq 1,00$: soal kategori mudah

Untuk mengetahui keefektifan soal dalam membedakan peserta tes kelompok atas dan kelompok bawah, maka dilakukan analisis daya pembeda. Daya pembeda butir soal berfungsi untuk menentukan dapat tidaknya suatu butir soal membedakan kelompok peserta tes yang berkemampuan tinggi dan berkemampuan rendah. Daya pembeda dapat dihitung dengan mencari nilai korelasi Biserial, secara matematis nilainya dinyatakan:

$$r_{bis} = \left(\frac{M_p - M_t}{S_T} \right) \left(\frac{p}{y} \right)$$

(Nonoh Siti Aminah, 2012: 9)

Keterangan:

- p : proporsi *testee* yang menjawab benar pada item soal
- q : Proporsi *testee* yang menjawab salah pada item soal (1 – p)
- M_p : mean skor tes dari peserta tes yang menjawab benar
- M_t : mean skor total
- S_T : varians total
- y : ordinat p dalam distribusi normal

Tabel 1. Klasifikasi Daya Beda

| Nilai p | Klasifikasi |
|-------------------------|---------------------|
| $p \leq 0,19$ | <i>Poor</i> |
| $0,20 \leq p \leq 0,40$ | <i>Satisfactory</i> |
| $0,40 \leq p \leq 0,70$ | <i>Good</i> |
| $0,70 \leq p \leq 1,00$ | <i>Excellent</i> |
| Bertanda negatif | ditolak |

Sebagai tindak lanjut dari peng-analisisan mengenai daya pembeda item, item soal dikelompokkan dengan dasar berikut :

1. Butir – butir soal yang sudah memiliki daya pembeda item yang baik (*satisfactory*, *good*, dan *excellent*) hendaknya dimasukkan dalam buku bank soal tes hasil belajar. Butir – butir item tersebut dapat dipakai.
2. Butir – butir item yang daya pembedanya masih rendah (*poor*), ada dua kemungkinan tindak lanjut, yaitu :
 - a) Ditelusuri untuk kemudian di-perbaiki, dan setelah diperbaiki dapat diajukan lagi dalam tes hasil belajar yang akan datang.
 - b) Dibuang dan tidak digunakan. (Anas Sudijono, 1995 :389).

Alternatif jawaban yang merupakan pengecoh (distraktor) yang baik harus memiliki koefisien korelasi yang negatif dan tinggi (Saifudin Azwar, 2002:141). Karena hal ini mengindikasikan bahwa pengecoh (distraktor) dipilih oleh siswa dari kelompok bawah di samping itu Fernandes (1984) menyimpulkan pengecoh dapat berfungsi baik jika paling sedikit dipilih oleh 2% peserta tes (Elvin Yusliana Ekawati, 2010:336).

Keputusan suatu item soal layak digunakan, perlu direvisi atau ditolak didasarkan pada kriteria keputusan untuk penilaian item soal oleh Elvin Yuslana Ekawati (2010) sebagai berikut

1. Item soal diterima, apabila karakteristik item soal memenuhi semua kriteria. Item soal yang terlalu sukar atau mudah, tetapi memiliki daya beda dan distribusi pengecoh item yang memenuhi kriteria, butir soal tersebut dapat diterima atau dipilih.
2. Item soal direvisi, apabila salah satu atau lebih dari ketiga kriteria karakteristik item soal tidak diterima.
3. Item soal ditolak, jika item soal memiliki karakteristik yang tidak memenuhi semua kriteria (2010:336).

METODE PENELITIAN

Penelitian dilaksanakan di SMA Negeri 1 Girimarto kelas X tahun Pelajaran 2011/2012. Penelitian dilaksanakan pada bulan Januari-Juni 2012 subjek penelitian adalah siswa kelas X.1, X.2, X.3, X.4, dan X.5.

Metode penelitian yang digunakan merupakan metode penelitian dasar. “Penelitian dasar adalah suatu penelitian yang bertujuan untuk mengembangkan teori ilmiah atau prinsip dasar suatu disiplin ilmu yang lebih baik daripada hanya memecahkan persoalan praktis” (Nonoh Siti A, 2002:1).

Analisis data dilakukan secara kuantitatif dan kualitatif. Analisis data secara kualitatif dilakukan untuk menganalisis butir soal yang telah disusun secara deskriptif, Penelaahan dilakukan meliputi aspek materi, konstruksi dan bahasa. Dalam penelaahan soal digunakan lembar telaah untuk mempermudah prosedur pelaksanaan. Adapun analisis kuantitatif yang dilakukan meliputi analisis validitas, tingkat kesukaran item soal, daya pembeda dan efektifitas disktraktor.

HASIL DAN PEMBAHASAN

Intrumen tes yang disusun adalah intrumen tes formatif tengah semester yang terdiri dari 20 indikator. Pada tahap awal dari 20 indikator tersebut disusun dua paket soal paket A dan B yang masing-masing terdiri dari 50 butir soal. Bentuk tes yang disusun adalah tes pilihan ganda (*multiple choice*) dengan lima pilihan jawaban. Sebelum instrumen tersebut diujikan di lapangan, instrumen perlu ditelaah secara kualitatif. Telaah dilakukan dari aspek materi, konstruksi dan bahasa. Dalam penelitian ini telaah dilakukan oleh dosen pembimbing dan guru mata pelajaran fisika. Hasil telaah kualitatif disajikan pada tabel 2.

Tabel 2 Rangkuman Telaah Diskriptif Soal yang Tidak Memenuhi Kriteria

| Aspek yang ditelaah | Paket A | | Paket B | |
|---------------------|--|--------|---|--------|
| | Nomor Soal | Jumlah | Nomor soal | Jumlah |
| Materi | 14, 28, 33 | 4 | 44 | 1 |
| Konstruksi | 1, 2, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 19, 20, 21, 22, 23, 26, 32, 35, 36, 37, 41, 42, 45, 47, 48, 49 | 28 | 1, 2, 5, 6, 8, 9, 10, 12, 13, 14, 17, 18, 19, 22, 31, 32, 33, 34, 34, 35, 36, 42, 48, 49, | 24 |

| | | | | |
|--------|--|----|--|----|
| Bahasa | 8, 11, 12, 16, 17, 18, 21, 22, 26, 27, 29, 31, 36, 37, 38, 39, 40, 41, 42, 46, 50 | 21 | 15, 24, 31, 36, 38, 39, 40, 41, 42, 46, 47, 50 | 12 |
|--------|--|----|--|----|

Aspek materi kesalahan umumnya terjadi karena pengecoh tidak homogen dan logis. Aspek konstruksi kesalahan terjadi karena gambar tidak diberi nomor dan justru diberi petunjuk “di atas” atau “di bawah”, penggunaan kata yang kurang efektif sehingga pernyataan soal terkesan berbelit-belit dan kurang singkat dan jelas serta terdapat pilihan jawaban “semua benar”. Untuk aspek bahasa kesalahan yang terjadi, yaitu: penggunaan pemilihan kata yang kurang tepat dan penulisan pernyataan soal tidak sesuai dengan kaidah penulisan dalam bahasa Indonesia (EYD).

Setelah dilakukan telaah kualitatif langkah selanjutnya adalah analisis secara kuantitatif. Analisis menggunakan *MicroCat Iteman Versi 3.00* dimana program ini secara otomatis akan memberikan nilai reliabilitas soal, daya beda, taraf kesukaran dan keefektifan pengecoh.

Tabel 3 Hasil Analisis Tingkat Kesukaran Soal paket A dan B

| Klasifikasi | Paket A | | Paket B | |
|-------------------------------------|--|-------------|--|-------------|
| | Nomor soal | Jumlah | Nomor soal | Jumlah |
| Mudah ($P > 0,7$) | 2, 4, 49 | 3 (6%) | - | 0 (0%) |
| Sedang ($0,3 \geq P \geq 0,7$) | 1, 3, 5, 6, 8, 9, 12, 13, 14, 16, 19, 24, 26, 27, 35, 39, 42, 43, 44, 46, 48 | 21 (42%) | 1, 2, 5, 6, 7, 8, 11, 13, 15, 16, 20, 23, 24, 25, 26, 27, 28, 35, 40, 41, 45, 49, 50 | 23 (46%) |
| Sukar ($P < 0,3$) | 7, 10, 11, 15, 17, 18, 20, 21, 22, 23, 25, 28, 29, 30, 31, 32, 33, 34, 36, 37, 38, 40, 41, 45, 47, 50 | 26 (54%) | 3, 4, 9, 10, 12, 14, 17, 18, 19, 21, 22, 29, 30, 31, 32, 33, 34, 36, 37, 38, 39, 42, 43, 44, 46, 47, 48 | 27 (54%) |

Berdasarkan hasil analisis pada Tabel 3, soal yang mempunyai daya baeda baik (soal kategori sedang) untuk soal paket A terdapat 21 butir, dan untuk soal paket B sebanyak 23 butir, sementara untuk soal kategori mudah untuk soal paket A terdapat 3 butir, dan kategori sukar untuk soal Paket A terdapat 26 butir dan paket B 27 butir.

Tabel 4 Hasil Daya Pembeda Soal Paket A dan B

| Klasifikasi | Paket A | | Paket B | | Keputusan |
|--------------------------------------|---|-------------|---|-------------|---|
| | Nomor soal | Jumlah | Nomor soal | Jumlah | |
| <i>Excellent</i> (0,70 – 1,00) | 10, 47 | 2 (4%) | 4, 19, 21, 32, 34, 37, 42, 44, 46, 47 | 10 (20%) | Diterima A (60%) B (70 %) |
| <i>Good</i> (0,40 – 0,70) | 3, 4, 5, 8, 11, 12, 24, 25, 26, 32, 40, 44, 46 | 13 (26%) | 3, 6, 9, 12, 14, 25, 31, 43, 48 | 9 (18%) | |
| <i>Satisfactory</i> (0,20 – 0,40) | 1, 7, 9, 13, 16, 20, 21, 23, 27, 28, 29, 30, 39, 42, 43 | 15 (30%) | 2, 5, 7, 10, 13, 15, 16, 22, 27, 28, 29, 36, 38, 40, 41, 50 | 16 (32%) | |
| <i>Poor</i> (0,00 – 0,20) | 2, 6, 14, 17, 33, 35, 37, 38, 45, 48 | 10 (20%) | 1, 11, 17, 18, 20, 30, 35, 45, 49 | 9 (18%) | Direvisi/ diteliti A (40%) B (30 %) |
| Jelek sekali ($< 0,00$) | 15, 18, 19, 22, 31, 34, 36, 41, 49, 50 | 10 (20%) | 8, 23, 24, 26, 33, 39 | 6 (12%) | |

Daya beda masuk dalam katagori diterima jika nilai korelasi biserial minimal bernilai 0,20 sehingga untuk soal paket A yang memenuhi kriteria baik sejumlah 30 butir dan peket B 35 butir, sementara daya beda yang masuk dalam katagori tidak baik untuk soal paket A sejumlah 20 butir dan paket B sejumlah 15 butir.

Tabel 5 Katagori Item Soal Berdasarkan Distribusi Pengecoh Item

| Pengecoh berfungsi | Paket A | | Paket B | | Keterangan |
|--------------------|---|-------------|---|-------------|------------|
| | Nomor | Jml | Nomor | Jml | |
| 4 | 3, 4, 7, 10, 11, 12, 13, 16, 17, 18, 19, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32, 33, 35, 36, 39, 40, 42, 46, 48 | 30 (60%) | 1, 2, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 33, 34, 35, 36, 38, 40, 41, 42, 43, 44, 45, 46, 47 | 40 (80%) | Baik |
| 3 | 5, 6, 8, 9, 14, 15, 24, 34, 38, 41, 43, 44, 45, 47 | 14 (28%) | 3, 5, 10, 29, 32, 37, 39, 48, 49, 50 | 10 (20%) | Tidak baik |
| 2 | 1, 37, 49, 50 | 4 (8%) | - | - | Tidak baik |
| 1 | 2, 20 | 2 (4%) | - | - | Tidak baik |

Distribusi pengecoh dilihat dari *proporsi endorsing* atau proporsi yang menjawab pengecoh. Pengecoh dikatakan berfungsi baik jika minimal dipilih oleh 2% sedangkan jika kurang dari 2% maka masuk dalam katagori tidak baik. Pada Tabel 5 disajikan katagori soal berdasarkan distribusi pengecoh. Dimana diperoleh hasil 30 butir soal paket A dan 40 butir soal paket B berkatagori baik.

Tabel 6 Rangkuman Keputusan Uji Coba I

| Katagori | Soal Paket A | | Soal Paket A | |
|----------|--|------------------|--|------------------|
| | Nomor Soal | Jumlah | Nomor Soal | Jumlah |
| Diterima | 3, 7, 10, 11, 12, 13, 16, 20, 21, 23, 25, 26, 27, 28, 29, 30, 32, 39, 40, 42, 46 | 21 soal (42%) | 2, 4, 6, 7, 9, 12, 13, 14, 15, 16, 19, 21, 22, 25, 27, 28, 31, 34, 36, 40, 41, 42, 43, 44, 46, 47, | 26 soal (52%) |
| Direvisi | 1, 2, 4, 5, 6, 8, 9, 14, 17, 18, 19, 22, 24, 31, 33, 35, 36, 43, 44, 47, 48 | 21 soal (42%) | 1, 3, 5, 8, 10, 11, 17, 18, 20, 23, 24, 26, 29, 30, 32, 33, 35, 37, 38, 45, 48, 49, 50 | 23 soal (46%) |
| Ditolak | 15, 34, 37, 38, 41, 45, 49, 50 | 8 soal (16%) | 39 | 1 soal (2%) |

Hasil akhir keterimaan soal disajikan pada Tabel 6. Dimana untuk soal paket A 21 butir diterima, 21 butir direvisi dan 8 butir ditolak. Sedangkan untuk soal paket B 26 butir diterima, 23 butir direvisi dan 1 butir ditolak.

Setelah hasil Uji I dianalisis langkah selanjutnya adalah menyusun paket soal baru dari kedua peket soal sebelumnya, paket soal baru yang disusun terdiri dari 40 butir soal dengan lama waktu pengerjaan 90 menit. Tabel 7 menyajikan rincian butir yang akan digunakan untuk paket soal baru.

Tabel 7 Rangkuman Butir Soal untuk Uji Kedua

| Kriteria | Soal Paket A | | Soal Paket B | |
|--------------------|--|--------|--|--------|
| | Nomor soal | Jumlah | Nomor soal | Jumlah |
| Langsung digunakan | 3, 7, 10, 11, 13, 16, 26, 27, 28, 29, 32, 40 | 12 | 6, 9, 14, 15, 19, 21, 22, 31, 34, 36, 42, 43, 44, 46, 47 | 15 |
| Direvisi | 8, 20, 24, 23, 33 | 5 | 2, 18, 37, 38, 41, 45, 48, 50 | 8 |

Uji II dilakukan pada 3 kelas yaitu X. 2 , X.4, dan X.5 SMA Negeri 1 Girimarto dengan jumlah peserta uji 84 siswa. Setelah dilakukan uji langkah selanjutnya menganalisis seperti analisis untuk Uji I.

Tabel 8 Hasil Analisis Tingkat Kesukaran

| Kategori | No soal | Jumlah |
|----------|--|---------------|
| Mudah | - | 0 (0%) |
| Sedang | 1, 2, 3, 5, 9, 11, 12, 20, 21, 23, 26, 30, 31, 33, 36, 37, 40 | 17 (42,5%) |
| Sukar | 4, 6, 7, 8, 10, 13, 14, 15, 16, 17, 18, 19, 22, 24, 25, 27, 28, 29, 32, 34, 35, 38, 39 | 23 (57,5%) |

Hasil analisis daya beda untuk Uji II disajikan pada Tabel 10 dimana diperoleh hasil 17 butir kategori sedang dan 23 butir kategori sukar.

Tabel 9 Hasil Analisis Daya Pembeda Soal

| Klasifikasi | Nomor soal | Jumlah | Keputusan |
|--------------------------------------|---|---------------|---|
| <i>Excellent</i> (0,70 – 1,00) | 7, 27, | 2 (5%) | Diterima 26 soal (65%) |
| <i>Good</i> (0,40 – 0,70) | 2, 5, 6, 12, 13, 15, 16, 17, 25, 26, 28, 29, 35, 39 | 14 (35%) | |
| <i>Satisfactory</i> (0,20 – 0,40) | 1, 3, 8, 9, 11, 14, 19, 20, 22, 40 | 10 (25%) | |
| <i>Poor</i> (0,00 – 0,20) | 4, 10, 21, 23, 24, 30, 31, 32, 34, 36, 37 | 11 (27,5%) | Direvisi/ diteliti 14 soal (35 %) |
| Jelek sekali (< 0,00) | 18, 33, 38 | 3 (7,5%) | |
| Jumlah | | 40 | |

Soal diterima jika nilai daya beda $\geq 0,20$ dan direvisi/ diteliti jika daya beda $< 0,20$. Dari hasil uji II diperoleh hasil 26 soal atau 65% soal diterima, 14 soal atau 35% direvisi/ diteliti. Sedangkan nilai rata-rata daya beda hasil analisis menggunakan program iteman adalah 0,338 nilai ini tergolong cukup baik.

Tabel 10 Kategori Item Soal Berdasarkan Distribusi Pengecoh Item

| Pengecoh berfungsi | Nomor Soal | jumlah | Keterangan |
|--------------------|--|-------------|------------|
| 4 | 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40 | 38 (95%) | Baik |
| 3 | 1, 12 | 2 (5%) | Tidak baik |
| 2 | - | - | Tidak baik |
| 1 | - | - | Tidak baik |

Tabel 10 memperlihatkan data jumlah pengecoh yang berfungsi untuk tiap butir soal dimana kriteria pengecoh dikatakan berfungsi jika dipilih minimal oleh 2% testee, sedangkan secara keseluruhan pengecoh dikatakan baik jika kesemua pengecoh (4 pengecoh) berfungsi apabila terdapat pengecoh yang tidak berfungsi maka pengecoh dikategorikan tidak baik dan perlu direvisi. Sebagian besar pengecoh dapat berfungsi baik, hanya terdapat 2 soal yang pengecohnya tidak bekerja dengan baik yaitu soal nomor 1 dimana pengecoh yang tidak berfungsi pengecoh E dan nomor 12 dimana pengecoh yang tidak berfungsi adalah pengecoh D.

Tabel 11 Rangkuman Keputusan Uji Coba II

| Kategori | No. Item | Jumlah |
|----------|---|----------|
| Diterima | 2, 3, 5, 6, 7, 8, 9, 11, 13, 14, 15, 16, 17, 19, 20, 22, 25, 26, 27, 28, 29, 35, 39, 40 | 24 (60%) |
| Direvisi | 1, 4, 10, 12, 18, 21, 23, 24, 30, 31, 32, 33, 34, 36, 37, 38 | 16 (40%) |
| Ditolak | - | - |
| Jumlah | | 40 soal |

Dari hasil uji II diperoleh hasil 24 soal atau 60% soal diterima, 16 soal atau 40% direvisi. Hasil ini sudah lebih baik dari hasil uji pertama dimana untuk uji pertama baik soal paket A maupun B diperoleh hasil 42% soal diterima.

KESIMPULAN

Dari hasil penyusunan instrumen tes formatif fisika kelas X tengah semester genap untuk materi prinsip kerja alat-alat optik serta suhu dan kalor. Bentuk tes berupa tes objektif pilihan ganda model lima pilihan berjumlah 40 butir dari 20 indikator, dan waktu pelaksanaan tes 90 menit. Dapat disimpulkan bahwa instrumen tes yang telah disusun memiliki karakteristik sebagai berikut:

1. Soal Paket A:
 - a. Taraf kesukaran: terdapat 6% katagori mudah, 42% katagori sedang dan 54% katagori sukar.
 - b. Daya beda: sangat baik (*excellent*) sebanyak 4%, baik (*good*) sebanyak 26%, cukup (*satisfactory*) sebanyak 30%, rendah (*poor*) sebanyak 20% dan sangat jelek 20%.
 - c. Keefektifan distraktor diperoleh hasil: 60% soal empat distraktor berfungsi dan 28% soal 3 distraktor berfungsi, 8% soal dua distraktor berfungsi dan 4% soal satu distraktor berfungsi.

Diperoleh hasil akhir 42% soal diterima dan 42% soal direvisi dan 16% ditolak
2. Soal Paket B:
 - a. Taraf kesukaran: terdapat 46% katagori sedang dan 54% katagori sukar.
 - b. Daya beda: sangat baik (*excellent*) sebanyak 20%, baik (*good*) sebanyak 18%, cukup (*satisfactory*) sebanyak 32%, rendah (*poor*) sebanyak 18% dan sangat jelek 12%.
 - c. Keefektifan distraktor diperoleh hasil 80% soal empat distraktor berfungsi dan 20% soal 3 distraktor berfungsi.

Diperoleh hasil akhir 52% soal diterima dan 46% soal direvisi dan 2% soal ditolak
3. Instrumen tes akhir yang merupakan gabungan dari soal paket A dan B yang berkualitas baik:
 - a. Taraf kesukaran: terdapat 42,5% katagori sedang dan 57,5% katagori sukar.

- b. Daya beda: sangat baik (*excellent*) sebanyak 5%, baik (*good*) sebanyak 35%, cukup (*satisfactory*) sebanyak 25%, rendah (*poor*) sebanyak 27,5% dan sangat jelek 7,5%.
 - c. Keefektifan distraktor diperoleh hasil 95% soal empat distraktor berfungsi dan 5% soal 3 distraktor berfungsi.
- Sehingga diperoleh hasil akhir 60% soal diterima dan 40% direvisi.

DAFTAR PUSTAKA

1. Aminah, N.S. (2012). *Dasar-Dasar pengukuran dan Statistika pada Pembelajaran Fisika*. Surakarta: UNS Press.
2. Arikunto, S. (2011). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
3. Azwar, S. (2002). *Tes Prestasi Fungsi Pengembangan Pengukuran Prestasi Belajar*. Yogyakarta: Pustaka Pelajar.
4. Cartono & Utari, T.S.G. (2006). *Penilaian Hasil Belajar*. Bandung. Prisma Press.
5. Depdiknas. (2008). *Panduan Analisis Butir soal*. Diperoleh 5 Januari 2012. Dari: http://teguhsasmitosdp1.files.wordpress.com/2010/05/analisis_soal1.pdf
6. Ekawati, E.Y. (2010). Analisis Kualitas Tes Fisika Semester I Kelas X SMA Ngaglik Kabupaten Sleman Menggunakan Pendekatan Teori Tes Klasik dan Teori Respon Butir. *Prosiding Seminar Nasional Fisika dan Pendidikan Science*. Hlm 335-342. Surakarta. FKIP Universitas Sebelas Maret.
7. Mardapi, D. (2008). *Teknik Penyusunan Instrumen tes dan Non Tes*. Yogyakarta: Mitra Cendikia Press.
8. Rammers, H. H., Gage, H.L. & Rummel, J.F. (1960). *A Practical Introduction To Measurement and Evaluation*. New York: Harper & Row Publisher.
9. Sudijono, A. (2005). *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada.
10. Thoha, C. (1994). *Teknik Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada.